



Frequently Asked Questions on EVAAS

Frequently Asked Questions: Spring 2013

1. [What test scores are used for EVAAS?](#)
2. [When will the 2012–13 EOG and EOC test data be available?](#)
3. [When will Standard 6 show up in EVAAS or populate into the NC Teacher Evaluation System?](#)
4. [How often are EVAAS data updated?](#)
5. [Will the 3rd Grade Pretest be reinstated and used to measure teacher growth in EVAAS?](#)
6. [Will Common Exams be included in school composite scores?](#)
7. [What assessments are included in determining Standard 8?](#)
8. [Why do we not have individual growth measures for students?](#)
9. [Will EVAAS use retest scores?](#)
10. [How do students who take **NCEXTEND2** populate into EVAAS?](#)
11. [How many days must a student be enrolled in a school to count towards teacher and school value-added scores? Why might a student with fewer days show up in diagnostic reports?](#)
12. [Do data from EVAAS fall under the Family Education Rights and Privacy Act \(FERPA\)?](#)
13. [How are online and credit recovery classes used in EVAAS?](#)
14. [How will Common Exams affect EVAAS and Standard 6? How will their predictors be determined?](#)
15. [How are the three years of data that are used to determine the Standard 6 rating used in determining a teacher's status?](#)
16. [How will teachers receive a growth index when they teach at a school where the students' stay is short term and there is continuous transition back to their home schools?](#)
17. [When teachers move in and out of tested subjects, how will their teacher growth values be handled?](#)
18. [Why is one standard error used for some reports while two standard errors are used for teacher effectiveness?](#)
19. [Why are students grouped into only three groups for the teacher reports?](#)
20. [How are teachers compared to each other?](#)
21. [What does it really mean to receive a teacher index that is greater than +/- 2?](#)
22. [What is the difference between the teacher estimate and teacher index and why?](#)
23. [How can a teacher meet expected growth in all classes/subjects but not meet expected growth overall?](#)
24. [What causes the discrepancies between the percent of teachers at each level of effectiveness between content areas \(for example, Biology and English Language Arts\)?](#)
25. [How can teacher value-added scores be positive when their students score below the proficiency cut score?](#)
26. [If your value-added data is very high one year, does that raise the target for expected/high growth for the next year? If your value-added data is very low one year, does that lower the expected/high target for the next year?](#)
27. [What are Normal Curve Equivalents?](#)
28. [How are NCEs different from Percentiles?](#)
29. [How are predictive scores and projected scores different?](#)
30. [Where can we see which other test scores make up a predicted score?](#)

31. [What is the minimum number of students needed to calculate a teacher value-added score?](#)
32. [What does growth look like for a student performing at the 97th percentile?](#)
33. [Are predicted/projected scores available to teachers BEFORE students take a course? If so, where can they be found?](#)
34. [How are CTE teachers compared to other teachers?](#)
35. [Will CTE tests be used to calculate the school overall composite?](#)
36. [How are CTE predictions and effectiveness determined?](#)
37. [Will CTE courses need to be verified in EVAAS for the students who are taking the WorkKeys? Will there need to be a specific percentage of students who must be tested like there is with EOCs?](#)

1. What test scores are used for EVAAS?

Growth Measures	Assessments
Standard 6 for Teachers	End of Grade Assessments (EOGs), End of Course Assessments (EOCs), Career and Technical Education Post-Assessments (CTE), and Common Exams
Standard 8 for Administrators	EOGs, EOCs, CTE, and Common Exams
School Accountability	EOGs and EOCs

2. When will the 2012–13 EOG and EOC test data be available?

Each year, the Department of Public Instruction's Accountability Services Division completes an extensive review of all test data to be sure that the data are accurate. North Carolina Department of Public Instruction (NCDPI) staff members complete this process before the State Board of Education (SBE) certifies the results of the State Testing Program, which typically happens at the August SBE meeting. After the SBE certifies the results, the NCDPI moves the data to SAS to begin value-added analyses. This process typically takes about two months.

The results of the 2012–13 State Testing Program will be delayed because of the transition to the Next Generation Assessments and the need to complete the standards setting process. The results, including both scale scores and achievement levels, will be presented to the SBE for certification in October 2013. The NCDPI is exploring methods for using the scale scores to produce value-added results for release in the early fall.

3. When will Standard 6 show up in EVAAS or populate into the NC Teacher Evaluation System?

On January 7, 2013, district and school EVAAS administrators were able to access teacher evaluation dashboards which display a teacher's ratings on Standards 1 – 6 of the NC Educator Evaluation System. Teachers across NC were able to access their individual teacher evaluation dashboards in EVAAS on February 1, 2013.

Ratings on Standard 6 were moved into the NC Educator Effectiveness System (NCEES) online tool during the month of January and released on February 15, 2013 to the teacher.

The timing of the release for data from the 2012–13 school year will depend on when scale scores are set for the Next Generation Assessments and Common Exams. Please see question 2 for more information.

4. How often are EVAAS data updated?

On the 15th of every month, the NCDPI pulls an enrollment file from NCWISE to send to SAS. This file determines which students are included in projection reports and other reports that use the current students attending a school.

Effective with the 2013–14 school year, the NCDPI will use data from PowerSchool to populate reports in EVAAS.

5. Will the 3rd Grade Pretest be reinstated and used to measure teacher growth in EVAAS?

The SBE has approved the administration of a Beginning-of-Grade 3 Test. The previous Grade 3 Pretest was an assessment of second grade knowledge and skills. The new Beginning-of-Grade 3 assessment will actually be a form of the third grade End-of-Grade assessment and will focus on reading only. This assessment may be used to measure teacher growth in EVAAS, but it will not be used in school accountability growth measures.

6. Will Common Exams be included in school composite scores?

No. According to current SBE policy, only results from the EOG and EOC assessments are included in the school-wide growth for school accountability purposes.

7. What assessments are included in determining Standard 8?

The principal's Standard 8 rating will consider any applicable administered EOCs, EOGs, Common Exams, and CTE assessments.

Because a principal's Standard 8 rating includes assessments that are not part of the State Testing Program, the rating may not be the same as the growth component of the school accountability model.

8. Why do we not have individual growth measures for students?

In the past, though schools used the ABC growth model to determine individual student projections and growth, the NCDPI did not encourage student-level growth reports. Since there is so much room for error in a single test score, student-level growth data is also not provided in EVAAS reporting.

EVAAS uses a longitudinally merged database of student scores in all subjects and can provide reliable estimates of the progress of groups of students on the basis of that information and projections regarding where students are likely to score on future tests, assuming average progress. However, the actual test score of an individual student on test day is only an approximation of his or her actual level of understanding. It is influenced by chance. Perhaps the student guesses and gets the question right; perhaps not. Perhaps the student reads a question or an answer wrongly, and gets the question right or wrong, as a result. Perhaps the noise level in the test room makes it difficult to concentrate and the student misses questions they know. Perhaps the student is just inordinately lucky on test day. Therefore, the current score of a single student, by itself is not a sufficiently reliable indicator upon which to base a determination of that student's progress.

This lack of reliability is why EVAAS requires a minimum number of students to produce a value-added score. When we aggregate many students from a teacher's classes, we can obtain an estimate of the teacher's effectiveness that is substantially more reliable than the individual student estimate. Additionally, SBE policy requires three years of value-added data before an educator receives a status of highly effective, effective, or in need of improvement.

9. Will EVAAS use retest scores?

During the 2012–13 school year, there will be no retests for assessments in the State Testing Program. In the past, EVAAS did not use retest scores to determine value-added scores; doing so would have been unfair to educators with large numbers of proficient students. A student who scored a level III or IV on an assessment could improve his or her scale score upon re-test, but this opportunity is not provided to the student.

10. How do students who take *NCEXTEND2* populate into EVAAS?

The NCDPI and SAS Institute are exploring the inclusion of *NCEXTEND2* test scores in EVAAS. The NCDPI is completing a study this year to link the results of *NCEXTEND2* assessments to the standard End-of-Grade and End-of-Course assessments. The results of the study will inform a final decision on this issue.

11. How many days must a student be enrolled in a school to count towards teacher and school value-added scores? Why might a student with fewer days show up in diagnostic reports?

A student must be enrolled for 140 days for a year-long class or 70 days for a semester-long class at the time of the test administration. If a student does not meet this requirement but takes the assessment, his or her data may still be included in a teacher's diagnostic report. This information allows teachers to use the information to

help guide their teaching. Even though the student shows up in the diagnostic report, the student's score will not be used for teacher or school value-added calculations.

12. Do data from EVAAS fall under the Family Education Rights and Privacy Act (FERPA)?

The State Board of Education policy governing EVAAS is TCS-C-021 (<http://sbepolicy.dpi.state.nc.us/policies/TCS-C-021.asp?pri=04&cat=C&pol=021&acr=TCS>), and the General Statutes referenced are General Statute §115C-319 (http://www.ncga.state.nc.us/EnactedLegislation/Statutes/HTML/BySection/Chapter_115C/GS_115C-319.html) and General Statute §115C-321 (http://www.ncga.state.nc.us/EnactedLegislation/Statutes/HTML/BySection/Chapter_115C/GS_115C-321.html).

13. How are online and credit recovery classes used in EVAAS?

Credit recovery class data are not included in EVAAS. Online course student data for courses provided by an external provider (such as NovaNet and OdysseyWare) are not reported on the EVAAS website. At this time, the NC Virtual Public School will have its teacher value-added reports for its instructors housed separately.

However, the results from EOGs and EOCs taken when a student is enrolled in an online or credit recovery class are used in school-wide accountability growth for READY accountability reporting. Value-added data used to determine Standard 6 ratings are only based on students taught in-person; the teacher's evaluation must only be based on instruction that is delivered in the traditional classroom setting. For credit recovery, the assessment results are not used because students have additional instructional time. Please see [question 9](#) for additional information.

14. How will Common Exams affect EVAAS and standard 6? How will their predictors be determined?

Data from the districts that administered the Common Exams at the end of first semester are being used to complete analyses. Once data from the second semester are used in conjunction with first semester data, the SAS Institute will produce value-added scores for teachers who administered Common Exams.

To measure the progress of students on the tests that are not administered in sequential years (like the courses covered by Common Exams), EVAAS uses the scores of all the students who took the test plus their past testing history in other subjects. EVAAS determines the relationship between past testing histories of students in other subjects and their scores on the current year's test to determine what is average progress for students with similar past testing histories. This step happens after students are tested.

In other words, by looking at the past testing histories of students across the state, we can predict where students should score if they have made the progress that was average for students like themselves academically. We can compare the score we predict to the score the student actually makes to determine whether the student made

average progress. We aggregate this information for all the students who took the test within a school, district, or classroom to provide the value added school, district, or teacher effect. So, conceptually, the value added effect is the difference between the progress that group of students made and the progress students similar to them made (on average) statewide.

15. How are the three years of data that are used to determine the Standard 6 rating used in determining a teacher's status?

To determine an educator's overall status, ratings from Standards 1 – 5 in the most recent year and a three-year average of data from Standard 6 are used. Only years in which a teacher has growth data for his or her own students will be used in the three-year average. A year in which the teacher's Standard 6 rating is based on school-wide growth does not count as one of the three years.

For school year 2011–12, a teacher with data on his or her own students received a Standard 6 rating based 70 percent on the teacher's individual value-added score and 30 percent on the school-wide growth score. The weighted average of these two values produced a score used to determine the rating. The 2011–12 Standard 6 rating is provided for illustration purposes only and is not included in the three-year average that will be used for accountability.

The SBE has determined that the Standard 6 rating will be based 100 percent on the teacher's individual growth score, beginning with the 2012–13 school year. The 2012–13 school year is the first of the three years that will contribute to the Standard 6 rating based on three years of data. This rating is then used in the status determination.

16. How will teachers receive a growth index when they teach at a school where the students' stay is short term and there is continuous transition back to their home schools?

Often, teachers in alternative schools will not have individual growth indexes. The SBE has an accountability model for alternative schools in which alternative schools select the indicators that they would like to use for accountability determinations. For school year 2011–12, the growth component of the alternate model was used as the school-wide growth component of the Standard 6 rating. The SBE will be finalizing policies around the Standard 6 rating for teachers in alternative schools, hospital schools, and other special schools.

17. When teachers move in and out of tested subjects, how will their teacher growth values be handled?

By the end of school year 2013–14, each teacher in North Carolina will have some measure of student growth. However, these measures may be difficult to combine given that some are produced through EVAAS, some are produced through a pre-post test model, and some are produced through the analysis of student work. The SBE will be deciding how to combine data gathered through different sources.

18. Why is one standard error used for some reports while two standard errors are used for teacher effectiveness?

Two standard errors are used for teacher value-added reports to ensure that all can have confidence in data used in teachers' evaluations. The diagnostic reports use only one standard error because their purpose is not for evaluation but for helping teachers focus their attention on students' needs.

19. Why are students grouped into only three groups for the teacher reports?

District and school reports use five groups (quintiles) because of the larger number of students that are present in each group: teacher reports only use three groups. There must be at least five students for any one group. There is higher probability that teachers will have groups without any information if the students are split into five groups.

20. How are teachers compared to each other?

The teacher's index is how the teacher's impact on instruction is calculated for all value-added purposes. The teacher estimate cannot be compared across teachers because it does not take the standard error into account. As a result, the Department of Public Instruction recommends using only the teacher index. The teacher index results from dividing the teacher estimate by its standard error. The index is comparable across teachers and is therefore the fair way to determine the effectiveness level category of does not meet expected growth, meets expected growth, or exceeds expected growth.

21. What does it really mean to receive a teacher index that is greater than +/- 2?

A teacher's impact on instruction is significantly better than the average teacher in the state when a teacher's index is greater than or equal to 2.0. A teacher's impact on instruction is significantly worse than the average teacher in the state when a teacher's index is less than -2.0.

22. What is the difference between the teacher estimate and teacher index and why?

The teacher estimate is the combined growth of all students who meet the minimum enrollment requirement described in [question 11](#). The index is calculated by dividing the teacher estimate by the standard error. This process standardizes the value-added score. Some teachers have a lot of students, such as a high school biology teacher with 30 students per class and 6 classes per year, while some teachers have a small group of students, such as a fourth-grade teacher with 20 students who take the English Language Arts and Math EOGs. Using the standard error to produce an index takes into account the fact that we have more data about the biology teacher and can, therefore, be more confident about the estimate of the biology teacher than we can be for the estimate of the fourth-grade teacher. When examining value-added reports,

users should always rely on the teacher index because it includes the standard error, and therefore allows us to fairly compare teachers across disciplines.

23. How can a teacher meet expected growth in all classes/subjects but not meet expected growth overall?

It is important to remember the number of students represented by a teacher's value-added score. Let's consider a fifth grade self-contained teacher with 25 students in his class. When EVAAS produces separate value-added scores for English Language Arts, Mathematics, and Science, each of those scores is based on the growth of only 25 students. However, when the teacher's composite is calculated for those three subjects, the value-added score is based on 75 students – each child has three test scores. Generally, when the number of students is larger, the standard error on the score is smaller. With more data, we can be more precise about the teacher's value-added score. Let's see how the math works:

Value-Added Score for Fifth Grade Teacher for Science:

Teacher Estimate = 10

Standard Error = 6 (based on 25 students)

Teacher Index: $(10/6) = 1.67$ (meets expected growth)

Value-Added Composite for Fifth Grade Teacher for English Language Arts, Mathematics, and Science:

Teacher Estimate = 10

Standard Error = 3 (based on 75 students)

Teacher Index = $(10/3) = 3.33$ (exceeds expected growth)

24. What causes the discrepancies between the percent of teachers at each level of effectiveness between content areas (for example, Biology and English Language Arts)?

There are larger numbers of students taking some state assessments compared to others. Additionally, evidence from across the country shows that English Language Arts teachers are less likely to have either an extremely positive or extremely negative impact on their students' growth relative to teachers of other content areas.

Researchers believe this trend exists because students' growth on English Language Arts assessments is influenced by outside forces (such as reading at home) more so than students' growth in other content areas (such as advanced math and science).

25. How can teacher value-added scores be positive when their students score below the proficiency cut score?

EVAAS looks at progress. It is possible for students to make excellent progress and still be below a standard if they start at a very low level. That is why EVAAS is fair to teachers regardless of whether they teach low- or high-performing students. EVAAS looks at how each teacher impacted the learning of his or her students.

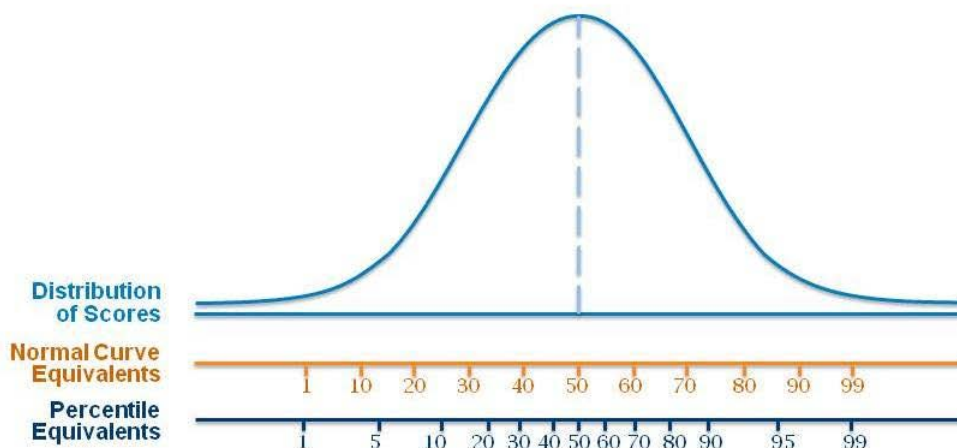
The SAS Institute runs the EVAAS models each year; each teacher is expected to make at least the average amount of growth for that grade and/or subject area with his or her students.

26. If one's value-added data is very high one year, does that raise the target for expected/high growth for the next year? If one's value-added data is very low one year, does that lower the expected/high target for the next year?

Value added scores are comparisons to the state average, which is recalculated every year. In the case of tested subjects, such as the EOC's, each teacher is compared to the average teacher in the state in that subject for that year. A teacher's performance from previous years does not raise the target for growth for the current year.

27. What are Normal Curve Equivalents?

Normal Curve Equivalents (NCEs) allow us to have a common language when we are comparing different tests and test versions even when the data does not fit a normal distribution or a bell curve. In essence, test scale scores are translated to State NCE's, which enable comparison of the achievement levels of groups of students from year to year. An NCE of 50 is the statewide average attainment level of students within a grade/subject/year.



28. How are NCEs different from Percentiles?

Percentiles show where a data point falls in relation to the entire set of data.

Example of Percentiles: Last year, Brian scored at the 97th percentile on the fifth grade science EOG

Interpretation: Brian scored better than 96 percent of students who took the fifth grade science EOG

Example of Normal Curve Equivalents:

A school's Estimated Mean NCE Score for fifth grade math might place it at the 40th percentile compared to other schools, but the normal curve equivalent is 45th percentile.

The normal curve equivalent looks at how far above or below the mean the value is.

29. How are predictive scores and projected scores different?

The Univariate Response Model (URM) is used for the EOGs in grades 5 and 8 science, the EOCs in high school, and the CTE Post-Assessments; URM is proposed for the Common Exams. The URM uses a student's prior test scores to **predict** a student's performance on a certain assessment.

To measure the progress of students on the tests that are not administered in sequential years (all tests EXCEPT Math and Reading), EVAAS uses the scores of all the students who took the test plus their past testing history in other subjects. After students are tested, EVAAS determines the relationship between past testing histories of children in other subjects and their scores on the current year's test to determine what is average progress for students with similar past testing histories.

In other words, EVAAS considers the past testing histories of students across the state and the relationship of that history to what the students actually scored. From that, it is possible to back up and look at their past testing history to predict where they should have scored, if they made the progress that was average for students like themselves, academically. EVAAS then compares the predicted score to the actual score to determine whether students made average progress. This information is aggregated for all the students who took that test within a school, district, or classroom to provide the value added school, district, or teacher effect. So, conceptually, the value added effect is the difference between the progress that group of students made and the progress students similar to them made, on average, statewide.

Projections are calculated in order to ascertain how future performance on assessments is likely to turn out for students with sufficient testing history, assuming they make the progress that is average for students like them, academically, statewide. At least three prior test scores (not necessarily in the tested subject) are used as predictors. The relationships that exist among scores for the most current set of test completers are used in order to compute projections. The Multivariate Response Model (MRM) is used for EOGs in grades 3–8 mathematics and English Language Arts. The EOGs in reading and math are standards that “build” on each other. As a result, the assessments have consecutive scale scores, and we use the MRM.

30. Where can we see which other test scores make up a predicted score?

In general, any test the student has taken in the past five years is considered in the prediction. There are exceptions, however, and these are the most common (Please keep in mind that there is a difference between the predictions used in the value added model for EOG Science, EOCs, and CTEs and the projections to future tests provided for students):

When predicting ACT, only EOG scores are used in the predictions for the school and district value added reports. This is because ACT does not show the effect of any specific class. For instance, progress in ACT Math is not attributable to ONLY Algebra I—the only high school math course tested. It is attributable to all math courses taught in the high school prior to students taking the ACT. So the value added reports for ACT are meant to show the progress students make in the high school across grades since 8th grade. We look at where those students were at the end of 8th grade and measure how much progress they make between then and when they take the ACT.

31. What is the minimum number of students needed to calculate a teacher value-added score?

Six students are needed for the Multivariate Response Model and ten students for the Univariate Response Model.

32. What does growth look like for a student performing at the 97th percentile?

North Carolina's expectation is that students will not lose ground academically from year to year relative to their statewide peers. North Carolina tests have sufficient "stretch" to measure the progress of students, whether they are well below grade level or well above it. For example, a seventh grader who scores at the 97th percentile should stay around 97% in eighth grade and not drop to the 50th percentile.

33. Are predicted/projected scores available to teachers BEFORE students take a course? If so, where can they be found?

Scores are not available, but projections for students' achievement levels are available from the Student History Reports, School and District Academic Preparedness Reports, Academic At-Risk Reports, and Student Search.

34. How are CTE teachers compared to other teachers?

CTE teachers are only compared to other CTE teachers teaching the same subject. There must be at least 15 teachers with a minimum of ten students each across the state teaching the same course for comparison. The minimum number 15 is subject to change, but sufficient data are always required in order to produce reliable, reflective value-added measures.

35. Will CTE tests be used to calculate the school overall composite?

According to current SBE policy, only results from the End-of-Grade and End-of-Course assessments are included in the school-wide accountability growth composite. Both the Common Exams and Career and Technical Education Post-Assessments are not part of the State Testing Program used to meet federal requirements under the Elementary and Secondary Education Act.

36. How are CTE predictions and effectiveness determined?

The value-added score for CTE exams is determined through the use of the predictive model, as are Algebra I, English II, and Biology. When value-added scores are calculated, CTE teachers are only compared to other CTE teachers in the state that teach the same class/course. So for example, Drafting 1 teachers are only compared to Drafting 1 teachers.

37. Will CTE courses need to be verified in EVAAS for the students who are taking the WorkKeys? Will there need to be a specific percentage of students who must be tested like there is with EOCs?

CTE teachers will verify their rosters in the EVAAS roster verification tool; however, districts may maintain their policies around exempting students from the CTE Post-Assessments due to scores on WorkKeys. The Department of Public Instruction has sent a memo to CTE directors, human resource directors, testing coordinators/directors, and curriculum directors with more information about this topic.

There is no 95 percent participation rule for the CTE Post-Assessments.